Asymptotic properties of Bernstein estimators on the simplex

Frédéric Ouimet[a,1,*]

[a] *California Institute of Technology, Pasadena, USA.*

Abstract

In this paper, we study various asymptotic properties (bias, variance, mean squared error, mean integrated squared error, asymptotic normality, uniform strong consistency) for Bernstein estimators of cumulative distribution functions and density functions on the d-dimensional simplex. Our results generalize the ones in Leblanc (2012a) and Babu *et al.* (2002), which treated the case $d = 1$, and significantly extend those found in Tenbusch (1994) for the density estimators when $d = 2$. The density estimator (or smoothed histogram) is closely related to the Dirichlet kernel estimator from Ouimet (2020a), and can also be used to analyze compositional data.

Keywords: Bernstein estimators, simplex, cumulative distribution function estimation, density estimation, mean squared error, asymptotic normality, uniform strong consistency
2010 MSC: Primary : 62G07 Secondary : 62G05, 62G20

1. Introduction

The d-dimensional simplex and its interior are defined by

$$\mathcal{S} := \big\{\boldsymbol{x} \in [0,1]^d : \|\boldsymbol{x}\|_1 \leq 1\big\} \quad \text{and} \quad \mathrm{Int}(\mathcal{S}) := \big\{\boldsymbol{x} \in (0,1)^d : \|\boldsymbol{x}\|_1 < 1\big\}, \tag{1.1}$$

where $\|\boldsymbol{x}\|_1 := \sum_{i=1}^d |x_i|$. For any cumulative distribution function F on $\mathcal{S}$, define the Bernstein polynomial of order m for F by

$$F_m^\star(\boldsymbol{x}) := \sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap m\mathcal{S}} F(\boldsymbol{k}/m) P_{\boldsymbol{k},m}(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{S}, \ m \in \mathbb{N}, \tag{1.2}$$

where the weights are the following probabilities from the Multinomial$(m, \boldsymbol{x})$ distribution :

$$P_{\boldsymbol{k},m}(\boldsymbol{x}) := \frac{m!}{(m - \|\boldsymbol{k}\|_1)! \prod_{i=1}^d k_i!} \cdot (1 - \|\boldsymbol{x}\|_1)^{m - \|\boldsymbol{k}\|_1} \prod_{i=1}^d x_i^{k_i}, \quad \boldsymbol{k} \in \mathbb{N}_0^d \cap m\mathcal{S}. \tag{1.3}$$

The Bernstein estimator of F, denoted $F_{n,m}^\star$, is the Bernstein polynomial of order m for the empirical cumulative distribution function $F_n(\boldsymbol{x}) := n^{-1} \sum_{j=1}^n \mathbb{1}_{(-\boldsymbol{\infty}, \boldsymbol{x}]}(\boldsymbol{X}_i)$, where the random variables $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_n$ are independent and F distributed. Precisely,

$$F_{n,m}^\star(\boldsymbol{x}) := \sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap m\mathcal{S}} F_n(\boldsymbol{k}/m) P_{\boldsymbol{k},m}(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{S}, \ m, n \in \mathbb{N}. \tag{1.4}$$

Similarly, if F has a density function f, we define the Bernstein density estimator of f by

$$\hat{f}_{n,m}(\boldsymbol{x}) := \sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}} \frac{m^d}{n} \sum_{i=1}^n \mathbb{1}_{(\frac{\boldsymbol{k}}{m}, \frac{\boldsymbol{k}+1}{m}]}(\boldsymbol{X}_i) P_{\boldsymbol{k},m-1}(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{S}, \ m, n \in \mathbb{N}, \tag{1.5}$$

where m^d is just a scaling factor, namely the inverse of the volume of the hypercube $\big(\frac{\boldsymbol{k}}{m}, \frac{\boldsymbol{k}+1}{m}\big]$.

[*] Corresponding author
Email address: ouimetfr@caltech.edu (Frédéric Ouimet)
[1] F. O. is supported by a postdoctoral fellowship from the NSERC (PDF) and the FRQNT (B3X supplement).

2. Results for the c.d.f. estimator $F^\star_{n,m}$

Except for Theorem 2.7, we assume the following everywhere in this section :

Assumption.

$$\bullet \quad \begin{array}{l} F \text{ is twice differentiable and its second order partial} \\ \text{derivatives are (uniformly) continuous on } \mathcal{S}. \end{array} \tag{2.1}$$

Proposition 2.1. *Under assumption (2.1), we have, uniformly for $\boldsymbol{x} \in \mathcal{S}$,*

$$F^\star_m(\boldsymbol{x}) = F(\boldsymbol{x}) + m^{-1} B(\boldsymbol{x}) + \mathrm{o}(m^{-1}), \tag{2.2}$$

as $m \to \infty$, where

$$B(\boldsymbol{x}) := \frac{1}{2} \sum_{i,j=1}^{d} \big(x_i \mathbb{1}_{\{i=j\}} - x_i x_j\big) \frac{\partial^2}{\partial x_i \partial x_j} F(\boldsymbol{x}). \tag{2.3}$$

Theorem 2.2 (Bias and variance)**.** *Under assumption (2.1), we have, for $\boldsymbol{x} \in \mathrm{Int}(\mathcal{S})$,*

$$\mathbb{B}\mathrm{ias}[F^\star_{n,m}(\boldsymbol{x})] = \mathbb{E}[F^\star_{n,m}(\boldsymbol{x})] - F(\boldsymbol{x}) = m^{-1} B(\boldsymbol{x}) + \mathrm{o}(m^{-1}), \tag{2.4}$$

$$\mathbb{V}\mathrm{ar}(F^\star_{n,m}(\boldsymbol{x})) = n^{-1}\sigma^2(\boldsymbol{x}) - m^{-1/2} n^{-1} V(\boldsymbol{x}) + \mathcal{O}_{\boldsymbol{x}}(m^{-1} n^{-1}), \tag{2.5}$$

as $m, n \to \infty$, where

$$\sigma^2(\boldsymbol{x}) := F(\boldsymbol{x})(1 - F(\boldsymbol{x})) \quad \text{and} \quad V(\boldsymbol{x}) := \sum_{i=1}^{d} \frac{\partial}{\partial x_i} F(\boldsymbol{x}) \sqrt{x_i(1-x_i)/\pi}. \tag{2.6}$$

Remark 2.3. *In Leblanc (2012a), the function $V(x)$ should be equal to $f(x)\sqrt{x(1-x)/\pi}$ instead of $f(x)\sqrt{2x(1-x)/\pi}$. The error is explained in the appendix and the estimates can easily be verified numerically. The same error also appears in the statements of Belalia (2016), since the proofs relied on the same estimates as Leblanc.*

Corollary 2.4 (Mean squared error)**.** *Under assumption (2.1), we have, for $\boldsymbol{x} \in \mathrm{Int}(\mathcal{S})$,*

$$\begin{aligned} \mathrm{MSE}(F^\star_{n,m}(\boldsymbol{x})) &= n^{-1}\sigma^2(\boldsymbol{x}) - n^{-1} m^{-1/2} V(\boldsymbol{x}) + m^{-2} B^2(\boldsymbol{x}) \\ &\quad + \mathcal{O}_{\boldsymbol{x}}(n^{-1} m^{-1}) + \mathrm{o}(m^{-2}). \end{aligned} \tag{2.7}$$

In particular, if $V(\boldsymbol{x}) \cdot B(\boldsymbol{x}) \neq 0$, the asymptotically optimal choice of m, with respect to MSE, is

$$m_{\mathrm{opt}} = n^{2/3} \left[\frac{4 B^2(\boldsymbol{x})}{V(\boldsymbol{x})}\right]^{2/3}, \tag{2.8}$$

in which case

$$\mathrm{MSE}[F^\star_{n,m_{\mathrm{opt}}}(\boldsymbol{x})] = n^{-1}\sigma^2(\boldsymbol{x}) - n^{-4/3} \frac{3}{4} \left[\frac{V^4(\boldsymbol{x})}{4 B^2(\boldsymbol{x})}\right]^{1/3} + \mathrm{o}_{\boldsymbol{x}}(n^{-4/3}). \tag{2.9}$$

Theorem 2.5 (Mean integrated squared error)**.** *Under assumption (2.1), we have*

$$\begin{aligned} \mathrm{MISE}[F^\star_{n,m}] &= n^{-1} \int_{\mathcal{S}} \sigma^2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} - n^{-1} m^{-1/2} \int_{\mathcal{S}} V(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} + m^{-2} \int_{\mathcal{S}} B^2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ &\quad + \mathrm{o}(n^{-1} m^{-1/2}) + \mathrm{o}(m^{-2}). \end{aligned} \tag{2.10}$$

In particular, if $\int_{\mathcal{S}} B^2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} > 0$, the asymptotically optimal choice of m, with respect to MISE, is

$$m_{\mathrm{opt}} = n^{2/3} \left[\frac{4 \int_{\mathcal{S}} B^2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}}{\int_{\mathcal{S}} V(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}}\right]^{2/3}, \tag{2.11}$$

in which case

$$\mathrm{MISE}[F^\star_{n,m_{\mathrm{opt}}}] = n^{-1} \int_{\mathcal{S}} \sigma^2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} - n^{-4/3} \frac{3}{4} \left[\frac{\big(\int_{\mathcal{S}} V(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}\big)^4}{4 \int_{\mathcal{S}} B^2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}}\right]^{1/3} + \mathrm{o}(n^{-4/3}). \tag{2.12}$$

Theorem 2.6 (Asymptotic normality)**.** *Assume* (2.1)*. For $\boldsymbol{x} \in \mathrm{Int}(\mathcal{S})$ such that $0 < F(\boldsymbol{x}) < 1$, we have the following convergence in distribution :*

$$n^{1/2}(F^\star_{n,m}(\boldsymbol{x}) - F^\star_m(\boldsymbol{x})) \stackrel{\mathscr{D}}{\longrightarrow} \mathcal{N}(0, \sigma^2(\boldsymbol{x})), \quad \text{as } m, n \to \infty. \tag{2.13}$$

In particular, Proposition 2.1 implies

$$n^{1/2}(F^\star_{n,m}(\boldsymbol{x}) - F(\boldsymbol{x})) \stackrel{\mathscr{D}}{\longrightarrow} \mathcal{N}(0, \sigma^2(\boldsymbol{x})), \qquad \text{if } n^{1/2}m^{-1} \to 0, \tag{2.14}$$

$$n^{1/2}(F^\star_{n,m}(\boldsymbol{x}) - F(\boldsymbol{x})) \stackrel{\mathscr{D}}{\longrightarrow} \mathcal{N}(\lambda\, B(\boldsymbol{x}), \sigma^2(\boldsymbol{x})), \quad \text{if } n^{1/2}m^{-1} \to \lambda, \tag{2.15}$$

for any constant $\lambda > 0$.

For the next result, we use the notation $\|G\|_\infty := \sup_{\boldsymbol{x} \in \mathcal{S}} |G(\boldsymbol{x})|$ for any bounded function $G : \mathcal{S} \to \mathbb{R}$, and also

$$\alpha_n := (n^{-1} \log n)^{1/2} \quad \text{and} \quad \beta_{n,m} := \alpha_n \sqrt{\alpha_m}. \tag{2.16}$$

Theorem 2.7 (Uniform strong consistency)**.** *Let F be continuous on $\mathcal{S}$. Then, as $m, n \to \infty$,*

$$\|F^\star_{n,m} - F\|_\infty \longrightarrow 0 \quad a.s. \tag{2.17}$$

Assume further that F is differentiable and its partial derivatives are Lipschitz continuous on $\mathcal{S}$. Then, for all $m \geq 2$ such that $m^{-1} \leq \beta_{n,m} \leq \alpha_m$ (for example, $2n^{2/3}/\log n \leq m \leq n^2/\log n$ works), we have, as $m, n \to \infty$,

$$\|F^\star_{n,m} - F_n\|_\infty = \mathcal{O}(\beta_{n,m}) \quad a.s. \tag{2.18}$$

In particular, for $m = n$, we have $\|F^\star_{n,m} - F_n\|_\infty = \mathcal{O}(n^{-3/4}(\log n)^{3/4})$ a.s.

3. Results for the density estimator $\hat{f}_{n,m}$

For each result stated in this section, one of the following two assumptions will be used.

Assumptions.

- *The density f is Lipschitz continuous on $\mathcal{S}$.* (3.1)
- *f is twice differentiable and its second order partial derivatives are (uniformly) continuous on $\mathcal{S}$.* (3.2)

We denote the expectation of $\hat{f}_{n,m}(\boldsymbol{x})$ by

$$f_m(\boldsymbol{x}) := \mathbb{E}[\hat{f}_{n,m}(\boldsymbol{x})] = \sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}} m^d \int_{(\frac{\boldsymbol{k}}{m}, \frac{\boldsymbol{k}+1}{m}]} f(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}\, P_{\boldsymbol{k},m}(\boldsymbol{x}). \tag{3.3}$$

Proposition 3.1. *Under assumption* (3.2)*, we have, uniformly for $\boldsymbol{x} \in \mathcal{S}$,*

$$f_m(\boldsymbol{x}) = f(\boldsymbol{x}) + m^{-1} b(\boldsymbol{x}) + \mathrm{o}(m^{-1}), \tag{3.4}$$

as $m \to \infty$, where

$$b(\boldsymbol{x}) := \sum_{i=1}^d \left(\tfrac{1}{2} - x_i\right) \frac{\partial}{\partial x_i} f(\boldsymbol{x}) + \frac{1}{2} \sum_{i,j=1}^d \left(x_i \mathbb{1}_{\{i=j\}} - x_i x_j\right) \frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{x}). \tag{3.5}$$

Theorem 3.2 (Bias and variance)**.** *We have, for $\boldsymbol{x} \in \mathrm{Int}(\mathcal{S})$,*

$$\mathbb{B}\mathrm{ias}[\hat{f}_{n,m}(\boldsymbol{x})] = \mathbb{E}[\hat{f}_{n,m}(\boldsymbol{x})] - f(\boldsymbol{x}) = m^{-1} b(\boldsymbol{x}) + \mathrm{o}(m^{-1}), \quad \text{assuming (3.2)}, \tag{3.6}$$

$$\mathbb{V}\mathrm{ar}(\hat{f}_{n,m}(\boldsymbol{x})) = n^{-1} m^{d/2} \psi(\boldsymbol{x}) f(\boldsymbol{x}) + \mathcal{O}_{\boldsymbol{x}}(n^{-1} m^{d/2 - 1/2}), \quad \text{assuming (3.1)}, \tag{3.7}$$

as $m, n \to \infty$, where

$$\psi(\boldsymbol{x}) := \left[(4\pi)^d x_1 x_2 \dots x_d (1 - \|\boldsymbol{x}\|_1)\right]^{-1/2}. \tag{3.8}$$

Corollary 3.3 (Mean squared error)**.** *Under assumption* (3.2)*, we have, for* $\boldsymbol{x} \in \mathrm{Int}(\mathcal{S})$*,*

$$\mathrm{MSE}(\hat{f}_{n,m}(\boldsymbol{x})) = n^{-1} m^{d/2} \psi(\boldsymbol{x}) f(\boldsymbol{x}) + m^{-2} b^2(\boldsymbol{x}) + \mathcal{O}_{\boldsymbol{x}}(n^{-1} m^{d/2-1/2}) + \mathrm{o}(m^{-2}). \tag{3.9}$$

In particular, if $f(\boldsymbol{x}) \cdot b(\boldsymbol{x}) \neq 0$*, the asymptotically optimal choice of* m*, with respect to* MSE*, is*

$$m_{\mathrm{opt}} = n^{2/(d+4)} \left[\frac{4}{d} \cdot \frac{b^2(\boldsymbol{x})}{\psi(\boldsymbol{x}) f(\boldsymbol{x})} \right]^{2/(d+4)}, \quad \textit{with} \tag{3.10}$$

$$\mathrm{MSE}[\hat{f}_{n,m_{\mathrm{opt}}}] = n^{-4/(d+4)} \left[\frac{\frac{4}{d} + 1}{\left(\frac{4}{d}\right)^{\frac{4}{d+4}}} \right] \frac{\big(\psi(\boldsymbol{x}) f(\boldsymbol{x})\big)^{4/(d+4)}}{\big(g^2(\boldsymbol{x})\big)^{-d/(d+4)}} + \mathrm{o}_{\boldsymbol{x}}(n^{-4/(d+4)}), \tag{3.11}$$

and, more generally, if $n^{2/(d+4)} m^{-1} \to \lambda$ *for some* $\lambda > 0$*, then*

$$\mathrm{MSE}[\hat{f}_{n,m}(\boldsymbol{x})] = n^{-4/(d+4)} \big[\lambda^{-d/2} \psi(\boldsymbol{x}) f(\boldsymbol{x}) + \lambda^2 b^2(\boldsymbol{x}) \big] + \mathrm{o}_{\boldsymbol{x}}(n^{-4/(d+4)}). \tag{3.12}$$

Theorem 3.4 (Mean integrated squared error)**.** *Under assumption* (3.2)*, we have*

$$\mathrm{MISE}[\hat{f}_{n,m}] = n^{-1} m^{d/2} \int_{\mathcal{S}} \psi(\boldsymbol{x}) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} + m^{-2} \int_{\mathcal{S}} b^2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} + \mathrm{o}(n^{-1} m^{d/2}) + \mathrm{o}(m^{-2}). \tag{3.13}$$

In particular, if $\int_{\mathcal{S}} b^2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} > 0$*, the asymptotically optimal choice of* m*, with respect to* MISE*, is*

$$m_{\mathrm{opt}} = n^{2/(d+4)} \left[\frac{4}{d} \cdot \frac{b^2(\boldsymbol{x})}{\psi(\boldsymbol{x}) f(\boldsymbol{x})} \right]^{2/(d+4)}, \quad \textit{with} \tag{3.14}$$

$$\mathrm{MISE}[\hat{f}_{n,m_{\mathrm{opt}}}] = n^{-4/(d+4)} \left[\frac{\frac{4}{d} + 1}{\left(\frac{4}{d}\right)^{\frac{4}{d+4}}} \right] \frac{\big(\int_{\mathcal{S}} \psi(\boldsymbol{x}) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}\big)^{4/(d+4)}}{\big(\int_{\mathcal{S}} g^2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}\big)^{-d/(d+4)}} + \mathrm{o}_{\boldsymbol{x}}(n^{-4/(d+4)}), \tag{3.15}$$

and, more generally, if $n^{2/(d+4)} m^{-1} \to \lambda$ *for some* $\lambda > 0$*, then*

$$\mathrm{MISE}[\hat{f}_{n,m}] = n^{-4/(d+4)} \left[\lambda^{-d/2} \int_{\mathcal{S}} \psi(\boldsymbol{x}) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} + \lambda^2 \int_{\mathcal{S}} b^2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \right] + \mathrm{o}(n^{-4/(d+4)}). \tag{3.16}$$

Theorem 3.5 (Uniform strong consistency)**.** *Assume* (3.1)*. If* $2 \leq m \leq n / \log n$ *as* $m, n \to \infty$*, then*

$$\begin{aligned} \|f_m - f\|_{\infty} &= \mathcal{O}(m^{-1/2}), \quad \textit{a.s.}, \\ \|\hat{f}_{n,m} - f\|_{\infty} &= \mathcal{O}(m^{d-1/2} \alpha_n) + \mathcal{O}(m^{-1/2}), \quad \textit{a.s.} \end{aligned} \tag{3.17}$$

In particular, if $m^{2d-1} = \mathrm{o}(n / \log n)$*, then* $\|\hat{f}_{n,m} - f\|_{\infty} \longrightarrow 0$ *a.s.*

Theorem 3.6 (Asymptotic normality)**.** *Assume* (3.1)*. Let* $\boldsymbol{x} \in \mathrm{Int}(\mathcal{S})$ *be such that* $f(\boldsymbol{x}) > 0$*. If* $n^{1/2} m^{-d/4} \to \infty$ *as* $m, n \to \infty$*, then*

$$n^{1/2} m^{-d/4} (\hat{f}_{n,m}(\boldsymbol{x}) - f_m(\boldsymbol{x})) \xrightarrow{\mathscr{D}} \mathcal{N}(0, \psi(\boldsymbol{x}) f(\boldsymbol{x})). \tag{3.18}$$

If we also have $n^{1/2} m^{-d/4-1/2} \to 0$ *as* $m, n \to \infty$*, then Theorem 3.5 implies*

$$n^{1/2} m^{-d/4} (\hat{f}_{n,m}(\boldsymbol{x}) - f(\boldsymbol{x})) \xrightarrow{\mathscr{D}} \mathcal{N}(0, \psi(\boldsymbol{x}) f(\boldsymbol{x})). \tag{3.19}$$

Independently of the above rates for n *and* m*, if we assume* (3.2) *instead and* $n^{2/(d+4)} m^{-1} \to \lambda$ *for some* $\lambda > 0$ *as* $m, n \to \infty$*, then Proposition 3.1 implies*

$$n^{2/(d+4)} (\hat{f}_{n,m}(\boldsymbol{x}) - f(\boldsymbol{x})) \xrightarrow{\mathscr{D}} \mathcal{N}(\lambda \, b(\boldsymbol{x}), \lambda^{-d/2} \psi(\boldsymbol{x}) f(\boldsymbol{x})). \tag{3.20}$$

Remark 3.7. *The rate of convergence for the* d*-dimensional kernel density estimator with i.i.d. data and bandwidth* h *is* $\mathcal{O}(n^{-1/2} h^{-d/2})$ *in Theorem 3.1.15 of Prakasa Rao (1983), whereas our estimator* $\hat{f}_{n,m}$ *converges at a rate of* $\mathcal{O}(n^{-1/2} m^{d/4})$*. Hence, the relation between the scaling factor* m *of* $\hat{f}_{n,m}$ *and the bandwidth* h *of other multivariate kernel smoothers is* $m \approx h^{-2}$*.*

4. Proof of the results for the c.d.f. estimator $F_{n,m}^\star$

Proof of Proposition 2.1. We generalize the proof of (Lorentz, 1986, Section 1.6.1), which treated the case $d = 1$. By the assumption (2.1), a second order mean value theorem yields

$$\begin{aligned} F(\boldsymbol{k}/m) - F(\boldsymbol{x}) &= \sum_{i=1}^d (k_i/m - x_i) \frac{\partial}{\partial x_i} F(\boldsymbol{x}) \\ &\quad + \frac{1}{2} \sum_{i,j=1}^d (k_i/m - x_i)(k_j/m - x_j) \frac{\partial^2}{\partial x_i \partial x_j} F(\boldsymbol{\xi_k}), \end{aligned} \tag{4.1}$$

for some random vector $\boldsymbol{\xi_k} \in \mathcal{S}$ on the line segment joining $\boldsymbol{k}/m$ and $\boldsymbol{x}$. Using the well-known identities

$$\sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap m\mathcal{S}} (k_i/m - x_i) P_{\boldsymbol{k},m}(\boldsymbol{x}) = 0, \tag{4.2}$$

and

$$\sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap m\mathcal{S}} (k_i/m - x_i)(k_j/m - x_j) P_{\boldsymbol{k},m}(\boldsymbol{x}) = \frac{1}{m}\big(x_i \mathbb{1}_{\{i=j\}} - x_i x_j\big), \tag{4.3}$$

we can multiply (4.1) by $P_{\boldsymbol{k},m}(\boldsymbol{x})$ and sum over $\boldsymbol{k} \in \mathbb{N}_0^d \cap m\mathcal{S}$ to obtain

$$\begin{aligned} &F_m^\star(\boldsymbol{x}) - F(\boldsymbol{x}) \\ &\quad = \sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap m\mathcal{S}} (F(\boldsymbol{k}/m) - F(\boldsymbol{x})) P_{\boldsymbol{k},m}(\boldsymbol{x}) \\ &\quad = \frac{1}{2m} \sum_{i,j=1}^d \big(x_i \mathbb{1}_{\{i=j\}} - x_i x_j\big) \frac{\partial^2}{\partial x_i \partial x_j} F(\boldsymbol{x}) \\ &\quad + \frac{1}{2} \sum_{i,j=1}^d \sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap m\mathcal{S}} (k_i/m - x_i)(k_j/m - x_j) \bigg(\frac{\partial^2}{\partial x_i \partial x_j} F(\boldsymbol{\xi_k}) - \frac{\partial^2}{\partial x_i \partial x_j} F(\boldsymbol{x})\bigg) P_{\boldsymbol{k},m}(\boldsymbol{x}). \end{aligned} \tag{4.4}$$

To conclude, we need to show that the last term is $\mathrm{o}(m^{-1})$. By the uniform continuity of the second order partial derivatives of F, we know that $\max_{1 \leq i,j \leq d} \| \frac{\partial^2}{\partial x_i \partial x_j} F \|_\infty \leq M_d$ for some $M_d > 0$, and we also know that, for all $\varepsilon > 0$, there exists $0 < \delta_{\varepsilon,d} \leq 1$ such that $\|\boldsymbol{y} - \boldsymbol{x}\|_1 \leq \delta_{\varepsilon,d}$ implies $\max_{1 \leq i,j \leq d} |\frac{\partial^2}{\partial x_i \partial x_j} F(\boldsymbol{y}) - \frac{\partial^2}{\partial x_i \partial x_j} F(\boldsymbol{x})| \leq \varepsilon$, uniformly for $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$. By considering the two cases $\|\boldsymbol{k}/m - \boldsymbol{x}\|_1 \leq \delta_{\varepsilon,d}$ and $\|\boldsymbol{k}/m - \boldsymbol{x}\|_1 > \delta_{\varepsilon,d}$, the last term in (4.4) is

$$\leq \frac{1}{2} \sum_{i,j=1}^d \Bigg[\varepsilon \sum_{\substack{\boldsymbol{k} \in \mathbb{N}_0^d \cap m\mathcal{S} \\ \|\boldsymbol{k}/m - \boldsymbol{x}\|_1 \leq \delta_{\varepsilon,d}}} |k_i/m - x_i||k_j/m - x_j| P_{\boldsymbol{k},m}(\boldsymbol{x}) \; + \; 2M_d \sum_{\ell=1}^d \sum_{\substack{\boldsymbol{k} \in \mathbb{N}_0^d \cap m\mathcal{S} \\ |k_\ell/m - x_\ell| > \delta_{\varepsilon,d}/d}} P_{\boldsymbol{k},m}(\boldsymbol{x})\Bigg]. \tag{4.5}$$

By Cauchy-Schwarz and the identity (4.3), the first term inside the bracket in (4.5) is

$$\leq \varepsilon \cdot \sqrt{m^{-1} x_i (1 - x_i)} \cdot \sqrt{m^{-1} x_j (1 - x_j)} \leq \frac{\varepsilon}{4m}. \tag{4.6}$$

By Bernstein's inequality (see e.g. Lemma A.1), the second term inside the bracket in (4.5) is

$$\leq 2M_d \cdot d \cdot 2 \exp\bigg(-\frac{(m\delta_{\varepsilon,d}/d)^2/2}{m \cdot 1 + \frac{1}{3} \cdot 1 \cdot (m\delta_{\varepsilon,d}/d)}\bigg) \leq 4d \, M_d \, e^{-\delta_{\varepsilon,d}^2 m/(4d^2)}. \tag{4.7}$$

If we take a sequence $\varepsilon = \varepsilon(m) \searrow 0$ such that $1 \geq \delta_{\varepsilon(m),d} \geq m^{-1/4}$, then (4.5) is $\mathrm{o}(m^{-1})$. $\square$

Proof of Theorem 2.2. The expression for the bias of $F^\star_{n,m}(\boldsymbol{x})$ just follows from Proposition 2.1 and the fact that

$$\mathbb{E}\big[F^\star_{n,m}(\boldsymbol{x})\big] = F^\star_m(\boldsymbol{x}), \quad \text{for all } \boldsymbol{x} \in \mathcal{S}. \tag{4.8}$$

To estimate the variance of $F^\star_{n,m}(\boldsymbol{x})$, note that

$$F^\star_{n,m}(\boldsymbol{x}) - F^\star_m(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{N}_0^d\cap m\mathcal{S}} (F_n(\boldsymbol{k}/m) - F(\boldsymbol{k}/m))P_{\boldsymbol{k},m}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^n Z_{i,m}, \tag{4.9}$$

where

$$Z_{i,m} := \sum_{\boldsymbol{k}\in\mathbb{N}_0^d\cap m\mathcal{S}} \big(\mathbb{1}_{(-\boldsymbol{\infty},\frac{\boldsymbol{k}}{m}]}(\boldsymbol{X}_i) - F(\boldsymbol{k}/m)\big)P_{\boldsymbol{k},m}(\boldsymbol{x}), \quad 1 \leq i \leq n. \tag{4.10}$$

For every m, the random variables $Z_{1,m},\dots,Z_{n,m}$ are i.i.d. and centered, so that

$$\begin{aligned}\mathbb{V}\mathrm{ar}(F^\star_{n,m}(\boldsymbol{x})) &= n^{-1}\,\mathbb{E}[Z_{1,m}^2] \\ &= n^{-1}\Big\{\sum_{\boldsymbol{k},\boldsymbol{\ell}\in\mathbb{N}_0^d\cap m\mathcal{S}} F((\boldsymbol{k}\wedge\boldsymbol{\ell})/m)P_{\boldsymbol{k},m}(\boldsymbol{x})P_{\boldsymbol{\ell},m}(\boldsymbol{x}) - \big(F^\star_m(\boldsymbol{x})\big)^2\Big\}.\end{aligned} \tag{4.11}$$

Using the expansion in (4.1) and Proposition 2.1, the above is

$$= n^{-1}\cdot\left\{\begin{array}{l} F(\boldsymbol{x})(1-F(\boldsymbol{x})) + \mathcal{O}(m^{-1}) \\ +\sum_{i=1}^d \frac{\partial}{\partial x_i}F(\boldsymbol{x})\sum_{\boldsymbol{k},\boldsymbol{\ell}\in\mathbb{N}_0^d\cap m\mathcal{S}}((k_i\wedge\ell_i)/m - x_i)P_{\boldsymbol{k},m}(\boldsymbol{x})P_{\boldsymbol{\ell},m}(\boldsymbol{x}) \\ +\sum_{i,j=1}^d \mathcal{O}\Big(\sum_{\boldsymbol{k},\boldsymbol{\ell}\in\mathbb{N}_0^d\cap m\mathcal{S}} |k_i/m - x_i||k_j/m - x_j|P_{\boldsymbol{k},m}(\boldsymbol{x})P_{\boldsymbol{\ell},m}(\boldsymbol{x})\Big)\end{array}\right\}. \tag{4.12}$$

The double sum on the second line inside the braces is estimated in (A.10) of Lemma A.3. By Cauchy-Schwarz, the identity (4.3), and the fact that $\sum_{\boldsymbol{\ell}\in\mathbb{N}_0^d\cap m\mathcal{S}} P_{\boldsymbol{\ell},m}(\boldsymbol{x}) = 1$, the double sum inside the big $\mathcal{O}$ term is

$$\leq \max_{1\leq i\leq n}\sum_{\boldsymbol{k},\boldsymbol{\ell}\in\mathbb{N}_0^d\cap m\mathcal{S}} |k_i/m - x_i|^2 P_{\boldsymbol{k},m}(\boldsymbol{x})P_{\boldsymbol{\ell},m}(\boldsymbol{x}) \leq \frac{1}{m}\max_{1\leq i\leq n} x_i(1-x_i) \leq \frac{1}{4m}. \tag{4.13}$$

This ends the proof. ☐

Proof of Theorem 2.5. By (4.12), (4.13) and (2.4), we have

$$\begin{aligned}\mathrm{MISE}(F^\star_{n,m}) &= \int_{\mathcal{S}}\Big(\mathbb{V}\mathrm{ar}(F^\star_{n,m}(\boldsymbol{x}) + \mathbb{B}\mathrm{ias}[F^\star_{n,m}(\boldsymbol{x})]^2\Big)\mathrm{d}\boldsymbol{x} \\ &= n^{-1}\left[\begin{array}{l}\int_{\mathcal{S}} F(\boldsymbol{x})(1-F(\boldsymbol{x}))\mathrm{d}\boldsymbol{x} + \mathcal{O}(m^{-1}) \\ +\sum_{i=1}^d \int_{\mathcal{S}} \frac{\partial}{\partial x_i}F(\boldsymbol{x})\sum_{\boldsymbol{k},\boldsymbol{\ell}\in\mathbb{N}_0^d\cap m\mathcal{S}}((k_i\wedge\ell_i)/m - x_i)P_{\boldsymbol{k},m}(\boldsymbol{x})P_{\boldsymbol{\ell},m}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\end{array}\right] \\ &\quad + m^{-2}\int_{\mathcal{S}} B^2(\boldsymbol{x})\mathrm{d}\boldsymbol{x} + \mathrm{o}(m^{-2}).\end{aligned} \tag{4.14}$$

By the assumption (2.1), the partial derivatives $(\frac{\partial}{\partial x_i}F)_{i=1}^d$ are bounded on $\mathcal{S}$, so Lemma A.3 and the bounded convergence theorem imply

$$\begin{aligned}\mathrm{MISE}(F^\star_{n,m}) &= n^{-1}\int_{\mathcal{S}} F(\boldsymbol{x})(1-F(\boldsymbol{x}))\mathrm{d}\boldsymbol{x} - n^{-1}m^{-1/2}\int_{\mathcal{S}}\sum_{i=1}^d \frac{\partial}{\partial x_i}F(\boldsymbol{x})\sqrt{\frac{x_i(1-x_i)}{\pi}}\,\mathrm{d}\boldsymbol{x} \\ &\quad + m^{-2}\int_{\mathcal{S}} B^2(\boldsymbol{x})\mathrm{d}\boldsymbol{x} + \mathrm{o}(n^{-1}m^{-1/2}) + \mathrm{o}(m^{-2}).\end{aligned} \tag{4.15}$$

This ends the proof. ☐

Proof of Theorem 2.6. Recall from (4.9) that $F^{\star}_{n,m}(\boldsymbol{x}) - F^{\star}_m(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^n Z_{i,m}$ where the $Z_{i,m}$'s are i.i.d. and centered random variables. Therefore, it suffices to show the following Lindeberg condition for double arrays :[2] For every $\varepsilon > 0$,

$$s_m^{-2}\,\mathbb{E}\big[Z_{1,m}^2 \mathbb{1}_{\{|Z_{1,m}|>\varepsilon n^{1/2} s_m\}}\big] \longrightarrow 0, \quad \text{as } n \to \infty. \tag{4.16}$$

where $s_m^2 := \mathbb{E}[Z_{1,m}^2]$ and where $m = m(n) \to \infty$. But this follows from the fact that $|Z_{1,m}| \leq 2$ for all m, and $s_m = (n\mathbb{V}\mathrm{ar}(F^{\star}_{n,m}))^{1/2} \to \sigma(\boldsymbol{x})$ as $m \to \infty$ by Theorem 2.2. □

Before proving Theorem 2.7, we need the following lemma (it is an adaptation of Lemma 2.2 in Babu & Chaubey (2006)).

Lemma 4.1. *Let F be Lipschitz continuous on $\mathcal{S}$, and let*[3]

$$N_{\boldsymbol{x},m} := \Big\{\boldsymbol{k} \in \mathbb{N}_0^d \cap m\mathcal{S} : \max_{1\leq i\leq d}\big|\tfrac{k_i}{m} - x_i\big| \leq \alpha_m\Big\}. \tag{4.17}$$

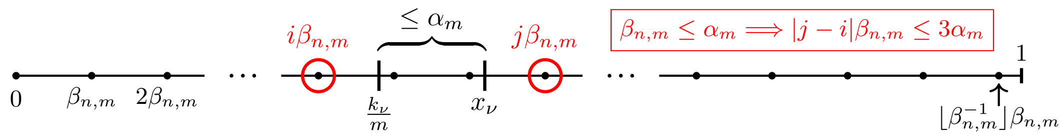
Then, for all $m \geq 2$ that satisfies $m^{-1} \leq \beta_{n,m} \leq \alpha_m$, we have, as $n \to \infty$,

$$\sup_{\boldsymbol{x}\in\mathrm{Int}(\mathcal{S})}\max_{\boldsymbol{k}\in N_{\boldsymbol{x},m}}\big|F_n(\boldsymbol{k}/m) - F(\boldsymbol{k}/m) - F_n(\boldsymbol{x}) + F(\boldsymbol{x})\big| = \mathcal{O}(\beta_{n,m}) \quad a.s. \tag{4.18}$$

Proof. For all $\boldsymbol{k} \in N_{\boldsymbol{x},m}$, we have

$$\begin{aligned}
&\big|F_n(\boldsymbol{k}/m) - F(\boldsymbol{k}/m) - F_n(\boldsymbol{x}) + F(\boldsymbol{x})\big| \\
&\leq \sum_{\nu=1}^d \left|\begin{array}{l} F_n\big(\frac{k_1}{m},\dots,\frac{k_{\nu-1}}{m},\frac{k_\nu}{m},x_{\nu+1},\dots,x_d\big) - F\big(\frac{k_1}{m},\dots,\frac{k_{\nu-1}}{m},\frac{k_\nu}{m},x_{\nu+1},\dots,x_d\big) \\ -F_n\big(\frac{k_1}{m},\dots,\frac{k_{\nu-1}}{m},x_\nu,x_{\nu+1},\dots,x_d\big) + F\big(\frac{k_1}{m},\dots,\frac{k_{\nu-1}}{m},x_\nu,x_{\nu+1},\dots,x_d\big)\end{array}\right| \\
&\leq \sum_{\nu=1}^d \max_{\substack{i,j\in\mathbb{N}_0 :\\ |i-j|\beta_{n,m}\leq 3\alpha_m}} \left|\begin{array}{l} F_n\big(\frac{k_1}{m},\dots,\frac{k_{\nu-1}}{m},j\beta_{n,m},x_{\nu+1},\dots,x_d\big) \\ -F\big(\frac{k_1}{m},\dots,\frac{k_{\nu-1}}{m},j\beta_{n,m},x_{\nu+1},\dots,x_d\big) \\ -F_n\big(\frac{k_1}{m},\dots,\frac{k_{\nu-1}}{m},i\beta_{n,m},x_{\nu+1},\dots,x_d\big) \\ +F\big(\frac{k_1}{m},\dots,\frac{k_{\nu-1}}{m},i\beta_{n,m},x_{\nu+1},\dots,x_d\big)\end{array}\right| + \mathcal{O}(\beta_{n,m}),
\end{aligned} \tag{4.19}$$

where the last inequality comes from our assumption that F is Lipschitz continuous.



For $\ell_\nu\beta_{n,m} < y_\nu \leq (\ell_\nu + 1)\beta_{n,m}$, $\nu = 1, 2, \dots, d$, and using the notation $\ell_\nu^+ = \ell_\nu + 1$, we have

$$\begin{aligned}
&\left|\begin{array}{l} F_n(y_1,\dots,y_{\nu-1},j\beta_{n,m},y_{\nu+1},\dots,y_d) - F(y_1,\dots,y_{\nu-1},j\beta_{n,m},y_{\nu+1},\dots,y_d) \\ -F_n(y_1,\dots,y_{\nu-1},i\beta_{n,m},y_{\nu+1},\dots,y_d) + F(y_1,\dots,y_{\nu-1},i\beta_{n,m},y_{\nu+1},\dots,y_d)\end{array}\right| \\
&\qquad\leq \left|\begin{array}{l} F_n(\ell_1^+\beta_{n,m},\dots,\ell_{\nu-1}^+\beta_{n,m},j\beta_{n,m},\ell_{\nu+1}^+\beta_{n,m},\dots,\ell_d^+\beta_{n,m}) \\ -F(\ell_1\beta_{n,m},\dots,\ell_{\nu-1}\beta_{n,m},j\beta_{n,m},\ell_{\nu+1}\beta_{n,m},\dots,\ell_d\beta_{n,m}) \\ -F_n(\ell_1\beta_{n,m},\dots,\ell_{\nu-1}\beta_{n,m},i\beta_{n,m},\ell_{\nu+1}\beta_{n,m},\dots,\ell_d\beta_{n,m}) \\ +F(\ell_1^+\beta_{n,m},\dots,\ell_{\nu-1}^+\beta_{n,m},i\beta_{n,m},\ell_{\nu+1}^+\beta_{n,m},\dots,\ell_d^+\beta_{n,m})\end{array}\right|
\end{aligned}$$

[2]See e.g. Section 1.9.3. in Serfling (1980).

[3]You can think of $N_{\boldsymbol{x},m}$ as the bulk of the Multinomial$(m,\boldsymbol{x})$ distribution; the contributions coming from outside the bulk are small for appropriate α_m's.

$$\leq \left| \begin{array}{l} F_n(\ell_1^+\beta_{n,m},\dots,\ell_{\nu-1}^+\beta_{n,m}, j\beta_{n,m}, \ell_{\nu+1}^+\beta_{n,m},\dots,\ell_d^+\beta_{n,m}) \\ -F_n(\ell_1\beta_{n,m},\dots,\ell_{\nu-1}\beta_{n,m}, i\beta_{n,m}, \ell_{\nu+1}\beta_{n,m},\dots,\ell_d\beta_{n,m}) \\ -F(\ell_1^+\beta_{n,m},\dots,\ell_{\nu-1}^+\beta_{n,m}, j\beta_{n,m}, \ell_{\nu+1}^+\beta_{n,m},\dots,\ell_d^+\beta_{n,m}) \\ +F(\ell_1\beta_{n,m},\dots,\ell_{\nu-1}\beta_{n,m}, i\beta_{n,m}, \ell_{\nu+1}\beta_{n,m},\dots,\ell_d\beta_{n,m}) \end{array} \right| + \mathcal{O}(\beta_{n,m})$$

$$\leq \sum_{\nu=1}^d D_{n,m,\nu} + \mathcal{O}(\beta_{n,m}), \tag{4.20}$$

where

$$D_{n,m,\nu} := \max_{\substack{i,j\in\mathbb{N}_0:\\ |i-j|\beta_{n,m}\leq 3\alpha_m}} \max_{\substack{0\leq k_p\leq 1+\lfloor\beta_{n,m}^{-1}\rfloor \\ p\in\{1,2,\dots,d\}\backslash\{\nu\}}} \left| \begin{array}{l} F_n(k_1\beta_{n,m},\dots,k_{\nu-1}\beta_{n,m}, j\beta_{n,m}, k_{\nu+1}\beta_{n,m},\dots,k_d\beta_{n,m}) \\ -F_n(k_1\beta_{n,m},\dots,k_{\nu-1}\beta_{n,m}, i\beta_{n,m}, k_{\nu+1}\beta_{n,m},\dots,k_d\beta_{n,m}) \\ -F(k_1\beta_{n,m},\dots,k_{\nu-1}\beta_{n,m}, j\beta_{n,m}, k_{\nu+1}\beta_{n,m},\dots,k_d\beta_{n,m}) \\ +F(k_1\beta_{n,m},\dots,k_{\nu-1}\beta_{n,m}, i\beta_{n,m}, k_{\nu+1}\beta_{n,m},\dots,k_d\beta_{n,m}) \end{array} \right|. \tag{4.21}$$

By (4.19), it follows that

$$\sup_{\boldsymbol{x}\in\mathrm{Int}(\mathcal{S})} \max_{\boldsymbol{k}\in N_{\boldsymbol{x},m}} \left|F_n(\boldsymbol{k}/m) - F(\boldsymbol{k}/m) - F_n(\boldsymbol{x}) + F(\boldsymbol{x})\right| \leq d \sum_{\nu=1}^d D_{n,m,\nu} + \mathcal{O}(\beta_{n,m}). \tag{4.22}$$

We want to apply a concentration bound on each $D_{n,m,\nu}$, $\nu = 1,2,\dots,d$. By Bernstein's inequality (see e.g. Lemma A.1), note that for any $\rho > 0$, any $y_1,\dots,y_{\nu-1},y_{\nu+1},\dots,y_d \in \mathbb{R}$ and any $i,j\in\mathbb{N}_0$ such that $|i-j|\,\beta_{n,m} \leq 3\alpha_m$, we have, assuming that $\beta_{n,m}\leq\alpha_m$,

$$\begin{aligned} &\mathbb{P}\left(\left| \begin{array}{l} F_n(y_1,\dots,y_{\nu-1}, j\beta_{n,m}, y_{\nu+1},\dots,y_d) \\ -F_n(y_1,\dots,y_{\nu-1}, i\beta_{n,m}, y_{\nu+1},\dots,y_d) \\ -F(y_1,\dots,y_{\nu-1}, j\beta_{n,m}, y_{\nu+1},\dots,y_d) \\ +F(y_1,\dots,y_{\nu-1}, i\beta_{n,m}, y_{\nu+1},\dots,y_d) \end{array} \right| \geq \rho\beta_{n,m} \right) \\ &\qquad \leq 2\exp\left(-\frac{\rho^2 n^2 \beta_{n,m}^2/2}{n\cdot C\cdot 3\alpha_m + \frac{1}{3}\cdot 1\cdot \rho n\beta_{n,m}} \right) \leq 2n^{-\rho^2/(8C)}, \end{aligned} \tag{4.23}$$

where $C \geq \rho$ is a Lipschitz constant for F. A union bound over i, j and the k_p's then yields

$$\mathbb{P}(D_{n,m,\nu} > \rho\beta_{n,m}) \leq (2 + \lfloor\beta_{n,m}^{-1}\rfloor)^{2+(d-1)} \cdot 2n^{-\rho^2/(8C)}, \quad 1\leq\nu\leq d. \tag{4.24}$$

Since $b_{n,m}^{-1} \leq n^2$ (indeed, our assumption $m^{-1} \leq b_{n,m}$ implies $b_{n,m}^{-1} \leq m$, and the second assumption $b_{n,m}\leq\alpha_m$ implies $m\leq n^2$), we can choose a constant $\rho = \rho(C,d) > 0$ large enough that the right-hand side of (4.24) is summable in n, in which case the Borel-Cantelli lemma implies $D_{n,m,\nu} = \mathcal{O}(\beta_{n,m})$ a.s. as $n\to\infty$. The conclusion follows from the bound in (4.22). ☐

Proof of Theorem 2.7. By the triangle inequality and $\sum_{\boldsymbol{k}\in\mathbb{N}_0^d\cap m\mathcal{S}} P_{\boldsymbol{k},m}(\boldsymbol{x}) = 1$, we have

$$\begin{aligned} \|F_{n,m}^\star - F\|_\infty &\leq \|F_{n,m}^\star - F\|_\infty + \|F_m^\star - F\|_\infty \\ &\leq \|F_n - F\|_\infty + \|F_m^\star - F\|_\infty. \end{aligned} \tag{4.25}$$

The first term on the last line goes to 0 by the Glivenko-Cantelli theorem, and the second term goes to 0 by the multidimensional Bernstein's theorem (i.e. a weak version of Proposition 2.1 where F is only assumed to be continuous on $\mathcal{S}$).[4]

[4] To be more precise, on the first line of (4.4), use the uniform continuity of F inside the bulk $N_{\boldsymbol{x},m}$ and a concentration bound to show that the contributions coming from outside the bulk are negligible. Alternatively, see Theorem 1.1.1 in Lorentz (1986).

For the remainder of the proof, we study the closeness between $F^{\star}_{n,m}$ and the empirical cumulative distribution function F_n. We assume that F is differentiable on $\mathcal{S}$ and its partial derivatives are Lipschitz continuous. By the triangle inequality,

$$
\begin{aligned}
\|F^{\star}_{n,m} - F_n\|_\infty \leq &\Big\| \sum_{\boldsymbol{k}\in N_{\boldsymbol{x},m}} (F_n(\boldsymbol{k}/m) - F(\boldsymbol{k}/m) - F_n(\,\cdot\,) + F(\,\cdot\,))P_{\boldsymbol{k},m}(\,\cdot\,)\Big\|_\infty \\
&+ \Big\| \sum_{\boldsymbol{k}\in(\mathbb{N}_0^d\cap m\mathcal{S})\backslash N_{\boldsymbol{x},m}} (F_n(\boldsymbol{k}/m) - F(\boldsymbol{k}/m) - F_n(\,\cdot\,) + F(\,\cdot\,))P_{\boldsymbol{k},m}(\,\cdot\,)\Big\|_\infty \\
&+ \Big\| \sum_{\boldsymbol{k}\in\mathbb{N}_0^d\cap m\mathcal{S}} (F(\boldsymbol{k}/m) - F(\,\cdot\,))P_{\boldsymbol{k},m}(\,\cdot\,)\Big\|_\infty .
\end{aligned}
\tag{4.26}
$$

The first norm is $\mathcal{O}(\beta_{n,m})$ by Lemma 4.1 (assuming $m^{-1} \leq \beta_{n,m} \leq \alpha_m$). If $X_i \sim \text{Binomial}(m, x_i)$, then a union bound, the fact that $\max_{\boldsymbol{k}} \|F_n(\boldsymbol{k}/m) - F(\,\cdot\,)\|_\infty \leq 1$, and Bernstein's inequality (see e.g. Lemma A.1), yield that the second norm in (4.26) is

$$
\begin{aligned}
\leq 2\cdot\max_{\boldsymbol{x}\in\mathcal{S}} \sum_{i=1}^d \mathbb{P}(|X_i - mx_i| \geq m\alpha_m) &\leq \max_{\boldsymbol{x}\in\mathcal{S}} 4\exp\Big(-\frac{m^2\alpha_m^2/2}{m\cdot x_i(1-x_i) + \frac{1}{3}\cdot 1\cdot m\alpha_m}\Big) \\
&\leq 4m^{-1} \leq 4\beta_{n,m}.
\end{aligned}
\tag{4.27}
$$

For the third norm in (4.26), the Lipschitz continuity of the partial derivatives $(\frac{\partial}{\partial x_i}F)_{i=1}^d$ implies that, uniformly for $\boldsymbol{x}\in\mathcal{S}$,

$$
F(\boldsymbol{k}/m) - F(\boldsymbol{x}) = \sum_{i=1}^d (k_i/m - x_i)\frac{\partial}{\partial x_i}F(\boldsymbol{x}) + \sum_{i,j=1}^d \mathcal{O}\big(|k_i/m - x_i||k_j/m - x_j|\big).
\tag{4.28}
$$

After multiplying (4.28) by $P_{\boldsymbol{k},m}(\boldsymbol{x})$, summing over $\boldsymbol{k}\in\mathbb{N}_0^d\cap m\mathcal{S}$ and applying the Cauchy-Schwarz inequality, the result is uniformly bounded by $\mathcal{O}(m^{-1})$ because of the identities (4.2) and (4.3). Since we assumed $m^{-1}\leq\beta_{n,m}$, this ends the proof. □

5. Proof of the results for the density estimator $\hat{f}_{n,m}$

Proof of Proposition 3.1. We follow the proof of Proposition 2.1. Using Taylor expansions for any $\boldsymbol{k}$ such that $\|\boldsymbol{k}/m - \boldsymbol{x}\|_1 = \mathrm{o}(1)$, we obtain

$$
\begin{aligned}
&m^d\int_{(\frac{\boldsymbol{k}}{m},\frac{\boldsymbol{k}+1}{m}]} f(\boldsymbol{y})\mathrm{d}\boldsymbol{y} - f(\boldsymbol{x}) \\
&\quad= f(\boldsymbol{k}/m) - f(\boldsymbol{x}) + \frac{1}{2m}\sum_{i=1}^d \frac{\partial}{\partial x_i}f(\boldsymbol{k}/m) + \mathcal{O}(m^{-2}) \\
&\quad= \frac{1}{m}\sum_{i=1}^d \frac{\partial}{\partial x_i}f(\boldsymbol{x})(k_i - mx_i) + \frac{1}{2m}\sum_{i=1}^d \frac{\partial}{\partial x_i}f(\boldsymbol{x}) + \mathrm{o}(m^{-1}) \\
&\qquad+ \frac{1}{2m^2}\sum_{i,j=1}^d \frac{\partial^2}{\partial x_i\partial x_j}f(\boldsymbol{x})(k_i - mx_i)(k_j - mx_j)(1+\mathrm{o}(1)) \\
&\quad= \frac{1}{m}\sum_{i=1}^d \frac{\partial}{\partial x_i}f(\boldsymbol{x})(k_i - (m-1)x_i) + \frac{1}{m}\sum_{i=1}^d \big(\tfrac{1}{2} - x_i\big)\frac{\partial}{\partial x_i}f(\boldsymbol{x}) \\
&\qquad+ \frac{1}{2m^2}\sum_{i,j=1}^d \frac{\partial^2}{\partial x_i\partial x_j}f(\boldsymbol{x})(k_i - mx_i)(k_j - mx_j)(1+\mathrm{o}(1)) + \mathrm{o}(m^{-1}).
\end{aligned}
\tag{5.1}
$$

If we multiply the last expression by $P_{\boldsymbol{k},m-1}(\boldsymbol{x})$ and sum over $\boldsymbol{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}$, then the identities (4.2) and (4.3) yield

$$
\begin{aligned}
&f_m(\boldsymbol{x}) - f(\boldsymbol{x}) \\
&\quad = 0 + \frac{1}{m} \sum_{i=1}^d \big(\tfrac{1}{2} - x_i\big) \frac{\partial}{\partial x_i} f(\boldsymbol{x}) + \frac{1}{2m} \sum_{i,j=1}^d \big(x_i \mathbb{1}_{\{i=j\}} - x_i x_j\big) \frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{x}) + \mathrm{o}(1),
\end{aligned} \tag{5.2}
$$

assuming that the o(1) rate in $\|\boldsymbol{k}/m - \boldsymbol{x}\|_1 = \mathrm{o}(1)$ decays slowly enough to 0 that the contributions coming from outside the bulk are negligible (exactly as we did in (4.7)). □

Proof of Theorem 3.2. The expression for the bias is a trivial consequence of Proposition 3.1 and the fact that $\mathbb{E}[\hat{f}_{n,m}(\boldsymbol{x})] = f_m(\boldsymbol{x})$. In order to compute the asymptotics of the variance, we only assume that f is Lipschitz continuous on $\mathcal{S}$. First, note that

$$
\hat{f}_{n,m}(\boldsymbol{x}) - f_m(\boldsymbol{x}) = \frac{m^d}{n} \sum_{i=1}^n Y_{i,m}, \tag{5.3}
$$

where

$$
Y_{i,m} := \sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}} \left[\mathbb{1}_{(\frac{\boldsymbol{k}}{m}, \frac{\boldsymbol{k}+1}{m}]}(\boldsymbol{X}_i) - \int_{(\frac{\boldsymbol{k}}{m}, \frac{\boldsymbol{k}+1}{m}]} f(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} \right] P_{\boldsymbol{k},m-1}(\boldsymbol{x}), \quad 1 \leq i \leq n. \tag{5.4}
$$

For every m, the random variables $Y_{1,m}, \dots, Y_{n,m}$ are i.i.d. and centered, so

$$
\mathbb{V}\mathrm{ar}(\hat{f}_{n,m}(\boldsymbol{x})) = n^{-1} m^{2d}\, \mathbb{E}[Y_{1,m}^2], \tag{5.5}
$$

and it is easy to see that

$$
\mathbb{E}[Y_{1,m}^2] = \sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}} \int_{(\frac{\boldsymbol{k}}{m}, \frac{\boldsymbol{k}+1}{m}]} f(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}\, P_{\boldsymbol{k},m-1}^2(\boldsymbol{x}) - \Big(m^{-d} f_m(\boldsymbol{x}) \Big)^2. \tag{5.6}
$$

The second term on the right-hand side of (5.6) is $\mathcal{O}(m^{-2d})$ since the Lipschitz continuity of f and the identity (4.3) imply that, uniformly for $\boldsymbol{x} \in \mathcal{S}$,

$$
f_m(\boldsymbol{x}) - f(\boldsymbol{x}) = \sum_{i=1}^d \mathcal{O}\bigg(\sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap (m-d)\mathcal{S}} |k_i/m - x_i| P_{\boldsymbol{k},m-1}(\boldsymbol{x}) \bigg) + \mathcal{O}(m^{-1}) = \mathcal{O}(m^{-1/2}). \tag{5.7}
$$

For the first term on the right-hand side of (5.6), the Lipschitz continuity of f implies,

$$
m^d \int_{(\frac{\boldsymbol{k}}{m}, \frac{\boldsymbol{k}+1}{m}]} f(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} = f(\boldsymbol{k}/m) + \mathcal{O}(m^{-1}) = f(\boldsymbol{x}) + \mathcal{O}(m^{-1}) + \sum_{i=1}^d \mathcal{O}\big(|k_i/m - x_i|\big), \tag{5.8}
$$

and by the Cauchy-Schwarz inequality, the identity (4.3) and (A.3) in Lemma A.2, we have, for all $i \in \{1, 2, \dots, d\}$,

$$
\begin{aligned}
&\sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}} |k_i/m - x_i|\, P_{\boldsymbol{k},m-1}^2(\boldsymbol{x}) \\
&\quad \leq \sqrt{\sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}} |k_i/m - x_i|^2 P_{\boldsymbol{k},m-1}(\boldsymbol{x})} \sqrt{\sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}} P_{\boldsymbol{k},m-1}^3(\boldsymbol{x})} = \mathcal{O}(m^{-1/2-d/2}).
\end{aligned} \tag{5.9}
$$

Putting (5.7), (5.8) and (5.9) together in (5.6) yields

$$
m^{3d/2}\, \mathbb{E}[Y_{1,m}^2] = (f(\boldsymbol{x}) + \mathcal{O}(m^{-1})) \Bigg[m^{d/2} \sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}} P_{\boldsymbol{k},m-1}^2(\boldsymbol{x}) \Bigg] + \mathcal{O}(m^{-1/2}). \tag{5.10}
$$

The result follows from (5.5) and (A.2) in Lemma A.2. □

Proof of Theorem 3.4. In Proposition 4.2 (a) of Ouimet (2018), it was shown, using the duplication formula for the Γ function and the ChuVandermonde convolution for binomial coefficients, that

$$m^{d/2} \int_{\mathcal{S}} \sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}} P_{\boldsymbol{k},m-1}^2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \int_{\mathcal{S}} \psi(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} + \mathcal{O}(m^{-1}). \tag{5.11}$$

Together with the almost-everywhere convergence in (A.2) of Lemma A.2, and the fact that f is bounded, Scheffé's lemma[5] implies

$$m^{d/2} \int_{\mathcal{S}} f(\boldsymbol{x}) \sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}} P_{\boldsymbol{k},m-1}^2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \int_{\mathcal{S}} \psi(\boldsymbol{x}) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} + \mathrm{o}(1). \tag{5.12}$$

Therefore, by (5.5), (5.10), (5.12) and (3.6), we have

$$\begin{aligned} \mathrm{MISE}(\hat{f}_{n,m}) &= \int_{\mathcal{S}} \Big(\mathbb{V}\mathrm{ar}(\hat{f}_{n,m}(\boldsymbol{x})) + \mathbb{B}\mathrm{ias}[\hat{f}_{n,m}(\boldsymbol{x})]^2 \Big) \mathrm{d}\boldsymbol{x} \\ &= n^{-1} m^{d/2} \int_{\mathcal{S}} \psi(\boldsymbol{x}) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} + m^{-2} \int_{\mathcal{S}} b^2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} + \mathrm{o}(n^{-1} m^{d/2}) + \mathrm{o}(m^{-2}). \end{aligned} \tag{5.13}$$

This ends the proof. □

Proof of Theorem 3.5. We have already shown that $\|f_m - f\|_\infty = \mathcal{O}(m^{-1/2})$ in (5.7). Next, we want to apply a concentration bound to control $\|\hat{f}_{n,m} - f_m\|_\infty$. Let

$$L_{n,m} := \max_{\boldsymbol{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}} \frac{1}{n} \sum_{i=1}^n \Big(\mathbb{1}_{(\frac{\boldsymbol{k}}{m}, \frac{\boldsymbol{k}+1}{m}]}(\boldsymbol{X}_i) - \int_{(\frac{\boldsymbol{k}}{m}, \frac{\boldsymbol{k}+1}{m}]} f(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} \Big). \tag{5.14}$$

By a union bound on $\boldsymbol{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}$ (there are at most m^d such points), and Bernstein's inequality (see e.g. Lemma A.1), we have, for all $\rho > 0$,

$$\begin{aligned} \mathbb{P}\Big(L_{n,m} > \rho m^{-1/2} \alpha_n \Big) &\leq m^d \cdot 2 \exp\Big(- \frac{\rho^2 n^2 m^{-1} \alpha_n^2 / 2}{n \cdot c \cdot m^{-1} + \frac{1}{3} \cdot 1 \cdot \rho n m^{-1/2} \alpha_n} \Big) \\ &\leq m^d \cdot n^{-\rho^2/(4c)}, \end{aligned} \tag{5.15}$$

where the second inequality assumes that $m \leq \frac{n}{\log n}$, and $c \geq \rho$ is a Lipschitz constant for f. If we choose $\rho = \rho(c,d) > 0$ large enough, then the right-hand side of (5.15) is summable in n and the Borel-Cantelli lemma implies $\|\hat{f}_{n,m} - f_m\|_\infty \leq m^d L_{n,m} = \mathcal{O}(m^{d-1/2} \alpha_n)$ a.s. as $n \to \infty$. □

Proof of Theorem 3.6. By (5.3), the asymptotic normality of $n^{1/2} m^{-d/4} (\hat{f}_{n,m}(\boldsymbol{x}) - f_m(\boldsymbol{x}))$ will follow if we verify the Lindeberg condition for double arrays :[6] For every $\varepsilon > 0$,

$$s_m^{-2} \, \mathbb{E}\big[|Y_{1,m}|^2 \mathbb{1}_{\{|Y_{1,m}| > \varepsilon n^{1/2} s_m\}} \big] \longrightarrow 0, \quad \text{as } n \to \infty, \tag{5.16}$$

where $s_m^2 := \mathbb{E}\big[|Y_{1,m}|^2 \big]$ and $m = m(n) \to \infty$. Clearly, from (5.4),

$$|Y_{1,m}| \leq \max_{\boldsymbol{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}} 2 \, P_{\boldsymbol{k},m}(\boldsymbol{x}) = \mathcal{O}(m^{-d/2}), \tag{5.17}$$

and we also know that $s_m = m^{-3d/4} \sqrt{\psi(\boldsymbol{x}) f(\boldsymbol{x})} (1 + \mathrm{o}_{\boldsymbol{x}}(1))$ when f is Lipschitz continuous, by the proof of Theorem 3.2, so

$$\frac{|Y_{i,m}|}{n^{1/2} s_m} = \mathcal{O}_{\boldsymbol{x}}(n^{-1/2} m^{-d/2} m^{3d/4}) = \mathcal{O}_{\boldsymbol{x}}(n^{-1/2} m^{d/4}) \longrightarrow 0, \tag{5.18}$$

[5] Scheffé's lemma can be found for example on page 55 of Williams (1991).

[6] See e.g. Section 1.9.3. in Serfling (1980).

whenever $n^{1/2}m^{-d/4} \to \infty$ as $m, n \to \infty$.[7] Under this condition, (5.16) holds and thus

$$n^{1/2}m^{-d/4}(\hat{f}_{n,m}(\boldsymbol{x}) - f_m(\boldsymbol{x})) = n^{1/2}m^{3d/4} \cdot \frac{1}{n}\sum_{i=1}^n Y_{i,m} \xrightarrow{\mathscr{D}} \mathcal{N}(0, f(\boldsymbol{x})\psi(\boldsymbol{x})). \tag{5.19}$$

This completes the proof of Theorem 3.6. □

A. Tools

The first lemma is a standard (but very useful) concentration bound, found for example in Corollary 2.11 of Boucheron *et al.* (2013).

Lemma A.1 (Bernstein's inequality)**.** *Let $X_1, X_2, \dots, X_n$ be a sequence of independent random variables such that $|X_i| \leq b < \infty$. Then, for all $t > 0$,*

$$\mathbb{P}\Big(\sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \geq t\Big) \leq 2\exp\Big(-\frac{t^2/2}{\sum_{i=1}^n \mathbb{E}[X_i^2] + \frac{1}{3}bt}\Big). \tag{A.1}$$

In the second lemma, we estimate sums of powers of multinomial probabilities. This is used in the proof of Theorem 3.2 and the proof of Theorem 3.4.

Lemma A.2. *For every $\boldsymbol{x} \in \mathrm{Int}(\mathcal{S})$, we have, as $r \to \infty$,*

$$r^{d/2} \sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap r\mathcal{S}} P_{\boldsymbol{k},r}^2(\boldsymbol{x}) = \big[(4\pi)^d x_1 x_2 \dots x_d (1 - \|\boldsymbol{x}\|_1)\big]^{-1/2} + \mathcal{O}_{\boldsymbol{x}}(r^{-1/2}), \tag{A.2}$$

$$r^{d} \sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap r\mathcal{S}} P_{\boldsymbol{k},r}^3(\boldsymbol{x}) = \big[(2\sqrt{3}\pi)^d x_1 x_2 \dots x_d (1 - \|\boldsymbol{x}\|_1)\big]^{-1} + \mathcal{O}_{\boldsymbol{x}}(r^{-1/2}). \tag{A.3}$$

Proof. It is well known that the covariance matrix of the multinomial distribution is $r\,\Sigma_{\boldsymbol{x}}$, where $\Sigma_{\boldsymbol{x}} = \mathrm{diag}(\boldsymbol{x}) - \boldsymbol{x}\boldsymbol{x}^\top$, see e.g. (Severini, 2005, p.377), and it is also known that

$$\det(\Sigma_{\boldsymbol{x}}) = x_1 x_2 \dots x_d (1 - \|\boldsymbol{x}\|_1), \tag{A.4}$$

see e.g. (Tanabe & Sagae, 1992, Theorem 1). Therefore, consider

$$\phi_{\Sigma_{\boldsymbol{x}}}(\boldsymbol{y}) := \frac{1}{\sqrt{(2\pi)^d \det(\Sigma_{\boldsymbol{x}})}} \cdot \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top \Sigma_{\boldsymbol{x}}^{-1} \boldsymbol{y}\Big), \quad \boldsymbol{y} \in \mathbb{R}^d, \tag{A.5}$$

the density of the multivariate normal $\mathcal{N}(\boldsymbol{0}, \Sigma_{\boldsymbol{x}})$. By a local limit theorem for the multinomial distribution (see e.g. Lemma 2 in Arenbaev (1976) or Theorem 2.1 in Ouimet (2020b)), we have

$$\begin{aligned}
r^{d/2} \sum_{\boldsymbol{k} \in \mathbb{N}_0^d \cap r\mathcal{S}} P_{\boldsymbol{k},r}^2(\boldsymbol{x}) &= \int_{\mathbb{R}^d} \phi_{\Sigma_{\boldsymbol{x}}}^2(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} + \mathcal{O}_{\boldsymbol{x}}(r^{-1/2}) \\
&= \frac{2^{-d/2}}{\sqrt{(2\pi)^d \det(\Sigma_{\boldsymbol{x}})}} \int_{\mathbb{R}^d} \phi_{\frac{1}{2}\Sigma_{\boldsymbol{x}}}(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} + \mathcal{O}_{\boldsymbol{x}}(r^{-1/2}) \\
&= \frac{2^{-d/2}}{\sqrt{(2\pi)^d \det(\Sigma_{\boldsymbol{x}})}} \cdot 1 + \mathcal{O}_{\boldsymbol{x}}(r^{-1/2}),
\end{aligned} \tag{A.6}$$

[7] The bound on $|Y_{1,m}|$ in the proof of Proposition 1 in Babu *et al.* (2002) is suboptimal when $d = 1$, this is why we get a slightly better rate in (5.18).

and

$$
\begin{aligned}
r^d \sum_{\boldsymbol{k}\in\mathbb{N}_0^d\cap r\mathcal{S}} P_{\boldsymbol{k},r}^3(\boldsymbol{x}) &= \int_{\mathbb{R}^d} \phi_{\Sigma_{\boldsymbol{x}}}^3(\boldsymbol{y})\mathrm{d}\boldsymbol{y} + \mathcal{O}_{\boldsymbol{x}}(r^{-1/2}) \\
&= \frac{3^{-d/2}}{(2\pi)^d\det(\Sigma_{\boldsymbol{x}})}\int_{\mathbb{R}^d} \phi_{\frac{1}{3}\Sigma_{\boldsymbol{x}}}(\boldsymbol{y})\mathrm{d}\boldsymbol{y} + \mathcal{O}_{\boldsymbol{x}}(r^{-1/2}) \\
&= \frac{3^{-d/2}}{(2\pi)^d\det(\Sigma_{\boldsymbol{x}})}\cdot 1 + \mathcal{O}_{\boldsymbol{x}}(r^{-1/2}). \qquad \text{(A.7)}
\end{aligned}
$$

This ends the proof. □

In the third lemma, we estimate another technical sum, needed in proof Theorem 2.2 and the proof of Theorem 2.5.

Lemma A.3. *For $i \in \{1,2,\dots,d\}$ and $r \in \mathbb{N}$, let*

$$
R_{i,r}(\boldsymbol{x}) := r^{1/2}\sum_{\boldsymbol{k},\boldsymbol{\ell}\in\mathbb{N}_0^d\cap r\mathcal{S}} ((k_i\wedge \ell_i)/r - x_i)P_{\boldsymbol{k},r}(\boldsymbol{x})P_{\boldsymbol{\ell},r}(\boldsymbol{x}), \quad \boldsymbol{x}\in\mathcal{S}. \qquad \text{(A.8)}
$$

Then,

$$
\sup_{1\le i\le d}\sup_{r\in\mathbb{N}}\sup_{\boldsymbol{x}\in\mathcal{S}} |R_{i,r}(\boldsymbol{x})| \le 1, \qquad \text{(A.9)}
$$

and for every $\boldsymbol{x}\in \mathrm{Int}(\mathcal{S})$, we have,

$$
R_{i,r}(\boldsymbol{x}) = -\sqrt{\frac{x_i(1-x_i)}{\pi}} + \mathcal{O}_{\boldsymbol{x}}(r^{-1/2}), \quad \text{as } r\to\infty. \qquad \text{(A.10)}
$$

Proof. By the Cauchy-Schwarz inequality and the identity (4.3), we have

$$
\begin{aligned}
|R_{i,r}(\boldsymbol{x})| &\le 2r^{1/2}\sum_{\boldsymbol{k}\in\mathbb{N}_0^d\cap r\mathcal{S}} |k_i/r - x_i|P_{\boldsymbol{k},r}(\boldsymbol{x}) \le 2r^{1/2}\sqrt{\sum_{\boldsymbol{k}\in\mathbb{N}_0^d\cap r\mathcal{S}} |k_i/r-x_i|^2 P_{\boldsymbol{k},r}(\boldsymbol{x})} \\
&\le 2r^{1/2}\cdot\sqrt{r^{-1}x_i(1-x_i)} \le 1. \qquad \text{(A.11)}
\end{aligned}
$$

For the second claim, we know that the marginal distributions of the multinomial are binomial, so if ϕ_{σ^2} denotes the density function of the $\mathcal{N}(0,\sigma^2)$ distribution, a standard local limit theorem for the binomial distribution (see e.g. Prokhorov (1953) or Theorem 2.1 in Ouimet (2020b)) and integration by parts show that

$$
\begin{aligned}
R_{i,r}(\boldsymbol{x}) &= 2\cdot x_i(1-x_i)\int_{-\infty}^{\infty}\frac{z}{x_i(1-x_i)}\,\phi_{x_i(1-x_i)}(z)\int_z^{\infty}\phi_{x_i(1-x_i)}(y)\mathrm{d}y\mathrm{d}z + \mathcal{O}_{\boldsymbol{x}}(r^{-1/2}) \\
&= 2\cdot x_i(1-x_i)\left[0 - \int_{-\infty}^{\infty}\phi_{x_i(1-x_i)}^2(z)\mathrm{d}z\right] + \mathcal{O}_{\boldsymbol{x}}(r^{-1/2}) \\
&= \frac{-2x_i(1-x_i)}{\sqrt{4\pi x_i(1-x_i)}}\int_{-\infty}^{\infty}\phi_{\frac{1}{2}x_i(1-x_i)}(z)\mathrm{d}z + \mathcal{O}_{\boldsymbol{x}}(r^{-1/2}) \\
&= -\sqrt{\frac{x_i(1-x_i)}{\pi}} + \mathcal{O}_{\boldsymbol{x}}(r^{-1/2}). \qquad \text{(A.12)}
\end{aligned}
$$

This ends the proof. □

Remark A.4. *The proof of (A.10) is much simpler here than the proof of Lemma 2 (iv) in Leblanc (2012a) ($d=1$), where a finely tuned continuity correction from Cressie (1978) was used to estimate the survival function instead of working with a local limit theorem directly. There is also a typo in Leblanc's paper, his function $\psi_2(x)$ should be equal to $[x(1-x)/(4\pi)]^{1/2}$ instead of $[x(1-x)/(2\pi)]^{1/2}$. As a consequence, his function $V(x)$ should be equal to $f(x)[x(1-x)/\pi]^{1/2}$ instead of $f(x)[2x(1-x)/\pi]^{1/2}$. The same error also affects the statements in Belalia (2016), since the proofs relied on the same estimates.*

References

Arenbaev, N. K. 1976. Asymptotic behavior of the multinomial distribution. *Teor. Veroyatnost. i Primenen.*, **21**(4), 826–831. MR0478288.

Babu, G. J., & Chaubey, Y. P. 2006. Smooth estimation of a distribution and density function on a hypercube using Bernstein polynomials for dependent random vectors. *Statist. Probab. Lett.*, **76**(9), 959–969. MR2270097.

Babu, G. J., Canty, A. J., & Chaubey, Y. P. 2002. Application of Bernstein polynomials for smooth estimation of a distribution and density function. *J. Statist. Plann. Inference*, **105**(2), 377–392. MR1910059.

Barrientos, A. F., Jara, A., & Quintana, F. A. 2015. Bayesian density estimation for compositional data using random Bernstein polynomials. *J. Statist. Plann. Inference*, **166**, 116–125. MR3390138.

Barrientos, A. F., Jara, A., & Quintana, F. A. 2017. Fully nonparametric regression for bounded data using dependent Bernstein polynomials. *J. Amer. Statist. Assoc.*, **112**(518).

Belalia, M. 2016. On the asymptotic properties of the Bernstein estimator of the multivariate distribution function. *Statist. Probab. Lett.*, **110**, 249–256. MR3474765.

Belalia, M., Bouezmarni, T., & Leblanc, A. 2017a. Smooth conditional distribution estimators using Bernstein polynomials. *Comput. Statist. Data Anal.*, **111**, 166–182. MR3630225.

Belalia, M., Bouezmarni, T., Lemyre, F. C., & Taamouti, A. 2017b. Testing independence based on Bernstein empirical copula and copula density. *J. Nonparametr. Stat.*, **29**(2), 346–380. MR3635017.

Belalia, M., Bouezmarni, T., & Leblanc, A. 2019. Bernstein conditional density estimation with application to conditional distribution and regression functions. *J. Korean Statist. Soc.*, **48**(3), 356–383. MR3983257.

Boucheron, S., Lugosi, G., & Massart, P. 2013. *Concentration inequalities. A nonasymptotic theory of independence*. Oxford University Press, Oxford. MR3185193.

Bouezmarni, T., & Rolin, J.-M. 2003. Consistency of the beta kernel density function estimator. *Canad. J. Statist.*, **31**(1), 89–98. MR1985506.

Bouezmarni, T., & Rolin, J.-M. 2007. Bernstein estimator for unbounded density function. *J. Nonparametr. Stat.*, **19**(3), 145–161. MR2351744.

Bouezmarni, T., & Scaillet, O. 2005. Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. *Econometric Theory*, **21**(2), 390–412. MR2179543.

Bouezmarni, T., M., Mesfioui, & Rolin, J. M. 2007. L_1-rate of convergence of smoothed histogram. *Statist. Probab. Lett.*, **77**(14), 1497–1504. MR2395599.

Bouezmarni, T., Rombouts, J. V. K., & Taamouti, A. 2010. Asymptotic properties of the Bernstein density copula estimator for α-mixing data. *J. Multivariate Anal.*, **101**(1), 1–10. MR2557614.

Bouezmarni, T., El Ghouch, A., & Taamouti, A. 2013. Bernstein estimator for unbounded copula densities. *Stat. Risk Model.*, **30**(4), 343–360. MR3143795.

Brown, B. M., & Chen, S. X. 1999. Beta-Bernstein smoothing for regression curves with compact support. *Scand. J. Statist.*, **26**(1), 47–59. MR1685301.

Chen, S. X. 1999. Beta kernel estimators for density functions. *Comput. Statist. Data Anal.*, **31**(2), 131–145. MR1718494.

Chen, S. X. 2000a. Beta kernel smoothers for regression curves. *Statist. Sinica*, **10**(1), 73–91. MR1742101.

Chen, S. X. 2000b. Probability density function estimation using gamma kernels. *Ann. Inst. Statist. Math*, **52**(3), 471–480. MR1794247.

Cressie, N. 1978. A finely tuned continuity correction. *Ann. Inst. Statist. Math.*, **30**(3), 435–442. MR538319.

Curtis, S. M., & Ghosh, S. K. 2011. A variable selection approach to monotonic regression with Bernstein polynomials. *J. Appl. Stat.*, **38**(5), 961–976. MR2782409.

Gawronski, W. 1985. Strong laws for density estimators of Bernstein type. *Period. Math. Hungar*, **16**(1), 23–43. MR0791719.

Gawronski, W., & Stadtmüller, U. 1980. On density estimation by means of Poisson's distribution. *Scand. J. Statist.*, **7**(2), 90–94. MR0574548.

Gawronski, W., & Stadtmüller, U. 1981. Smoothing histograms by means of lattice and continuous distributions. *Metrika*, **28**(3), 155–164. MR0638651.

Gawronski, W., & Stadtmüller, U. 1984. Linear combinations of iterated generalized Bernstein functions with an application to density estimation. *Acta Sci. Math.*, **47**(1-2), 205–221. MR0755576.

Ghosal, S. 2001. Convergence rates for density estimation with Bernstein polynomials. *Ann. Statist.*, **29**(5), 1264–1280. MR1873330.

Guan, Z. 2016. Efficient and robust density estimation using Bernstein type polynomials. *J. Nonparametr. Stat.*, **28**(2), 250–271. MR3488598.

Igarashi, G., & Kakizawa, Y. 2014. On improving convergence rate of Bernstein polynomial density estimator. *J. Nonparametr. Stat.*, **26**(1), 61–84. MR3174309.

Janssen, P., Swanepoel, J., & Veraverbeke, N. 2012. Large sample behavior of the Bernstein copula estimator. *J. Statist. Plann. Inference*, **142**(5), 1189–1197. MR2879763.

Janssen, P., Swanepoel, J., & Veraverbeke, N. 2014. A note on the asymptotic behavior of the Bernstein estimator of the copula density. *J. Multivariate Anal.*, **124**, 480–487. MR3147339.

Janssen, P., Swanepoel, J., & Veraverbeke, N. 2016. Bernstein estimation for a copula derivative with application to conditional distribution and regression functionals. *TEST*, **25**(2), 351–374. MR3493523.

Janssen, P., Swanepoel, J., & Veraverbeke, N. 2017. Smooth copula-based estimation of the conditional density function with a single covariate. *J. Multivariate Anal.*, **159**, 39–48. MR3668546.

Kakizawa, Y. 2004. Bernstein polynomial probability density estimation. *J. Nonparametr. Stat.*, **16**(5), 709–729. MR2068610.

Kakizawa, Y. 2006. Bernstein polynomial estimation of a spectral density. *J. Time Ser. Anal.*, **27**(2), 253–287. MR2235846.

Kakizawa, Y. 2011. A note on generalized Bernstein polynomial density estimators. *Stat. Methodol.*, **8**(2), 136–153. MR2769276.

Lange, K. 1995. Applications of the Dirichlet distribution to forensic match probabilities. *Genetica*, **96**(1-2), 107–117. doi:10.1007/BF01441156.

Leblanc, A. 2009. Chung-Smirnov property for Bernstein estimators of distribution functions. *J. Nonparametr. Stat.*, **21**(2), 133–142. MR2488150.

Leblanc, A. 2010. A bias-reduced approach to density estimation using Bernstein polynomials. *J. Nonparametr. Stat.*, **22**(3-4), 459–475. MR2662607.

Leblanc, A. 2012a. On estimating distribution functions using Bernstein polynomials. *Ann. Inst. Statist. Math.*, **64**(5), 919–943. MR2960952.

Leblanc, A. 2012b. On the boundary properties of Bernstein polynomial estimators of density and distribution functions. *J. Statist. Plann. Inference*, **142**(10), 2762–2778. MR2925964.

Lorentz, G. G. 1986. *Bernstein polynomials*. Second edition. Chelsea Publishing Co., New York. MR0864976.

Lu, L. 2015. On the uniform consistency of the Bernstein density estimator. *Statist. Probab. Lett.*, **107**, 52–61. MR3412755.

Ouimet, F. 2018. Complete monotonicity of multinomial probabilities and its application to Bernstein estimators on the simplex. *J. Math. Anal. Appl.*, **466**(2), 1609–1617. MR3825458.

Ouimet, F. 2019. *Extremes of log-correlated random fields and the Riemann-zeta function, and some asymptotic results for various estimators in statistics*. PhD thesis, Université de Montréal. `http://hdl.handle.net/1866/22667`.

Ouimet, F. 2020a. Density estimation using Dirichlet kernels. *Preprint*, 1–14. arXiv:2002.06956.

Ouimet, F. 2020b. A precise local limit theorem for the multinomial distribution. *Preprint*, 1–7. arXiv:2001.08512.

Petrone, S. 1999a. Bayesian density estimation using Bernstein polynomials. *Canad. J. Statist.*, **27**(1), 105–126. MR1703623.

Petrone, S. 1999b. Random Bernstein polynomials. *Scand. J. Statist.*, **26**(3), 373–393. MR1712051.

Petrone, S., & Wasserman, L. 2002. Consistency of Bernstein polynomial posteriors. *J. Roy. Statist. Soc. Ser. B*, **64**(1), 79–100. MR1881846.

Prakasa Rao, B. L. S. 1983. *Nonparametric functional estimation*. Probability and Mathematical Statistics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York. MR0740865.

Prakasa Rao, B. L. S. 2005. Estimation of distribution and density functions by generalized Bernstein polynomials. *Indian J. Pure Appl. Math.*, **36**(2), 63–88. MR2153833.

Prokhorov, Y. V. 1953. Asymptotic behavior of the binomial distribution. *Uspekhi Mat. Nauk*, **8**(3(55)), 135–142. MR56861.

Sancetta, A. 2007. Nonparametric estimation of distributions with given marginals via Bernstein-Kantorovich polynomials: L_1 and pointwise convergence theory. *J. Multivariate Anal.*, **98**(7), 1376–1390. MR2364124.

Sancetta, A., & Satchell, S. 2004. The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, **20**(3), 535–562. MR2061727.

Serfling, R. J. 1980. *Approximation theorems of mathematical statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York. MR0595165.

Severini, T. A. 2005. *Elements of distribution theory*. Cambridge Series in Statistical and Probabilistic Mathematics, vol. 17. Cambridge University Press, Cambridge. MR2168237.

Stadtmüller, U. 1983. Asymptotic distributions of smoothed histograms. *Metrika*, **30**(3), 145–158. MR0726014.

Stadtmüller, U. 1986. Asymptotic properties of nonparametric curve estimates. *Period. Math. Hungar.*, **17**(2), 83–108. MR0858109.

Tanabe, K., & Sagae, M. 1992. An exact Cholesky decomposition and the generalized inverse of the variance-covariance matrix of the multinomial distribution, with applications. *J. Roy. Statist. Soc. Ser. B*, **54**(1), 211–219. MR1157720.

Tenbusch, A. 1994. Two-dimensional Bernstein polynomial density estimators. *Metrika*, **41**(3-4), 233–253. MR1293514.

Tenbusch, A. 1997. Nonparametric curve estimation with Bernstein estimates. *Metrika*, **45**(1), 1–30. MR1437794.

Turnbull, B. C., & Ghosh, S. K. 2014. Unimodal density estimation using Bernstein polynomials. *Comput. Statist. Data Anal.*, **72**, 13–29. MR3139345.

Vitale, R. A. 1975. Bernstein polynomial approach to density function estimation. *Pages 87–99 of: Statistical Inference and Related Topics*. Academic Press, New York. MR0397977.

Williams, D. 1991. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge. MR1155402.